



Pettigrew, R. G. (2016). Accuracy, Risk, and the Principle of Indifference. *Philosophy and Phenomenological Research*, 92(1), 35-59.  
10.1111/phpr.12097

Peer reviewed version

Link to published version (if available):  
[10.1111/phpr.12097](http://10.1111/phpr.12097)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/pure/about/ebr-terms.html>

### Take down policy

Explore Bristol Research is a digital archive and the intention is that deposited content should not be removed. However, if you believe that this version of the work breaches copyright law please contact [open-access@bristol.ac.uk](mailto:open-access@bristol.ac.uk) and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline of the nature of the complaint

On receipt of your message the Open Access Team will immediately investigate your claim, make an initial judgement of the validity of the claim and, where appropriate, withdraw the item in question from public view.

# Accuracy, Risk, and the Principle of Indifference (forthcoming in *Philosophy and Phenomenological Research*)\*

Richard Pettigrew<sup>†</sup>

January 11, 2014

## 1 Introduction

In Bayesian epistemology, the *problem of the priors* is this: How should we set our credences (or degrees of belief) in the absence of evidence? That is, how should we set our *prior* or *initial* credences, the credences with which we begin our credal life? David Lewis liked to call an agent at the beginning of her credal journey a *superbaby*.<sup>1</sup> The problem of the priors asks for the norms that govern these superbabies.

The Principle of Indifference gives a very restrictive answer. It demands that such an agent divide her credences equally over all possibilities. That is, according to the Principle of Indifference, only one initial credence function is permissible, namely, the uniform distribution. In this paper, we offer a novel argument for the Principle of Indifference. I call it the *Argument from Accuracy*.

The Argument from Accuracy shares much in common with Jim Joyce's "nonpragmatic vindication" of Probabilism [Joyce, 1998]. Thus, in section 2, I introduce the Argument from Accuracy by first sketching a very restricted version of Joyce's argument for Probabilism; then, drawing on the framework introduced for that first sketch, I sketch an equally restricted version of my Argument from Accuracy. This will allow us to see the main

---

\*I would like to thank a number of people for helpful discussion of earlier versions of this paper: Lara Buchak, Teddy Groves, Jim Joyce, Jason Konek, Juergen Landes, Katie Steele, and Jon Williamson, as well as anonymous referees for this journal. I was supported by ERC Starting Grant EUT during my work on this paper.

<sup>†</sup>Department of Philosophy, University of Bristol, Cotham House, Bristol, UK. BS6 6JL.

<sup>1</sup>Alan Hájek reports that he first heard Lewis use the term in an undergraduate lecture at Princeton in the late 1980s [Hájek, ms].

philosophical ideas behind the argument so that we might compare it with rival arguments for the Principle of Indifference in section 3 and so that we might assess one of its central premises in section 4. In section 5, I lift the restrictions: I generalise the argument and then strengthen it by weakening the assumptions it makes. In section 6, I consider how the Principle of Indifference interacts with other norms that govern credences, such as Probabilism, Regularity, and the Principal Principle. I conclude in section 7 with a brief discussion of language dependence.

## 2 The Argument from Accuracy

In this section, we consider an agent who entertains only two propositions,  $A$  and  $B$ , which are exclusive and exhaustive. We represent such an agent's cognitive state at a given time in her credal life by her *credence function* at that time: this is the function  $c$  that takes each proposition about which she has an opinion—that is,  $A$  or  $B$ —and returns the real number that measures her *degree of belief* or *credence* in that proposition. By convention, we represent minimal credence by 0 and maximal credence by 1.

### 2.1 Joyce's argument for Probabilism

We will present Joyce's argument for the following norm, which governs such an agent:

(Prob<sub>2</sub>) At any time in an agent's credal life, if she entertains only  $A$  and  $B$  at that time, it ought to be the case that her credence function  $c$  at that time is such that

$$c(A) + c(B) = 1.$$

This amounts to the following pair of claims: as the agent's credence in one proposition rises, her credence in the other should fall by the same amount; and when the agent has maximal credence in one proposition, she should have minimal credence in the other. It is the strongest norm governing such an agent that follows from Probabilism.

How do we establish (Prob<sub>2</sub>)? Joyce offers the following argument: It is often said that the aim of full belief is truth. One way to make this precise is to say that the ideal doxastic state is that in which one believes every true proposition about which one has an opinion, and one disbelieves every false proposition about which one has an opinion. That is, the ideal doxastic state is the *omniscient doxastic state* (relative to the set of propositions about which one has an opinion). Thus, if one entertains only  $A$  and  $B$ , and  $A$  is true and  $B$  is false, then the omniscient doxastic state is that in

which one believes  $A$  and disbelieves  $B$ . We might then measure how good an agent's doxastic state is by its proximity to this omniscient state.<sup>2</sup>

Joyce's argument—as I will present it—is based on an analogous claim about credences. We say that the ideal credal state is that in which one assigns maximal credence (i.e. 1) to each true proposition about which one has an opinion and minimal credence (i.e. 0) to each false proposition about which one has an opinion. By analogy with the doxastic case, we might call this the *omniscient credal state* (relative to the set of propositions about which she has an opinion). Given a possible world  $w$ , we let  $v_w$  be the omniscient credence function at  $w$ . Thus:

- If  $A$  is true and  $B$  is false at  $w$ , we have:  $v_w(A) = 1; v_w(B) = 0$ .
- If  $A$  is false and  $B$  is true at  $w$ , we have:  $v_w(A) = 0; v_w(B) = 1$ .

We then measure how good an agent's credal state is by its proximity to the omniscient state. Following Joyce, we call this the *gradational accuracy* of the credal state (or, to abbreviate, its *accuracy*). To do this, we need a measure of distance between credence functions. Many different measures will do the job, but in this sketch of the argument I will focus on the most popular, namely, Squared Euclidean Distance.<sup>3</sup> Suppose  $c$  and  $c'$  are two credence functions defined for propositions  $A$  and  $B$  only. Then define the Squared Euclidean Distance between them as follows:

$$Q(c, c') := |c(A) - c'(A)|^2 + |c(B) - c'(B)|^2$$

That is, to obtain the distance between  $c$  and  $c'$ , we consider each proposition on which they are defined; we take the difference between the credences they assign to that proposition; we square that difference; and we sum the results. Thus, given a possible world  $w$ , the cognitive badness or disvalue of a credence function  $c$  at  $w$  is given by its inaccuracy; that is, the distance between  $c$  and  $v_w$ , namely,

$$Q(v_w, c) = |v_w(A) - c(A)|^2 + |v_w(B) - c(B)|^2$$

We call this the *Brier score* of  $c$  at  $w$ , and we write it  $B(c, w)$ . So, if we measure distance using Squared Euclidean Distance, then the cognitive value of  $c$  at  $w$  is the negative of the Brier score of  $c$  at  $w$ ; that is, it is  $-B(c, w)$ . Thus,  $B$  is a measure of inaccuracy;  $-B$  is a measure of accuracy.

With this measure of cognitive value in hand, Joyce argues for Probabilism by appealing to a standard norm of traditional decision theory:

(Dominance) Suppose  $\mathcal{O}$  is a set of options,  $\mathcal{W}$  is the set of possible worlds, and  $U$  is a measure of the value of the options in  $\mathcal{O}$  at the worlds in  $\mathcal{W}$ . Suppose  $o, o'$  in  $\mathcal{O}$ . Then we say that

<sup>2</sup>See [Easwaran, ms] for a fascinating description of the consequences of such an account of full beliefs.

<sup>3</sup> In fact, Joyce considers a broad range of distance measures; in Section 5.2, we do too.

- (a)  $o$  strongly  $U$ -dominates  $o'$  if  $U(o', w) < U(o, w)$  for all worlds  $w$  in  $\mathcal{W}$
- (b)  $o$  weakly  $U$ -dominates  $o'$  if  $U(o', w) \leq U(o, w)$  for all worlds  $w$  in  $\mathcal{W}$  and  $U(o', w) < U(o, w)$  for at least one world  $w$  in  $\mathcal{W}$ .

Now suppose  $o, o'$  in  $\mathcal{O}$  and

- (i)  $o$  strongly  $U$ -dominates  $o'$ ;
- (ii) There is no  $o''$  in  $\mathcal{O}$  that weakly  $U$ -dominates  $o$ .

Then  $o'$  is irrational.

Of course, in standard decision theory, the options are practical actions between which we wish to choose. But there is no reason why (Dominance) or any other decision-theoretic norm can only determine the irrationality of such options. They can equally be used to establish the irrationality of accepting a particular scientific theory or, as we will see, the irrationality of particular credal states. When they are put to use in the latter way, the options are the possible credal states and the measure of value is  $-B$ , the negative of the Brier score. Granted that, which credal states does (Dominance) rule out? As the following theorem shows, it is precisely those that violate (Prob<sub>2</sub>). This theorem is illustrated by Figure 2.1.

**Theorem 1** *For all credence functions  $c$ :*

- (I) *If  $c$  violates (Prob<sub>2</sub>), then there is a credence function  $c^*$  that satisfies (Prob<sub>2</sub>) such that  $c^*$  strongly Brier dominates  $c$ .*
- (II) *If  $c$  satisfies (Prob<sub>2</sub>), then there is no credence function  $c^*$  such  $c^*$  weakly Brier dominates  $c$ .*

This, then, is Joyce's argument for (Prob<sub>2</sub>):

- (1) The cognitive value of a credence function is given by its proximity to the ideal credence function:
  - (i) The ideal credence function at world  $w$  is the omniscient credence function at  $w$ , namely,  $v_w$ .
  - (ii) Distance is measured by the Squared Euclidean Distance.

Thus, the cognitive value of a credence function at a world is given by the negative of its Brier score at that world.

- (2) (Dominance)
- (3) Theorem 1

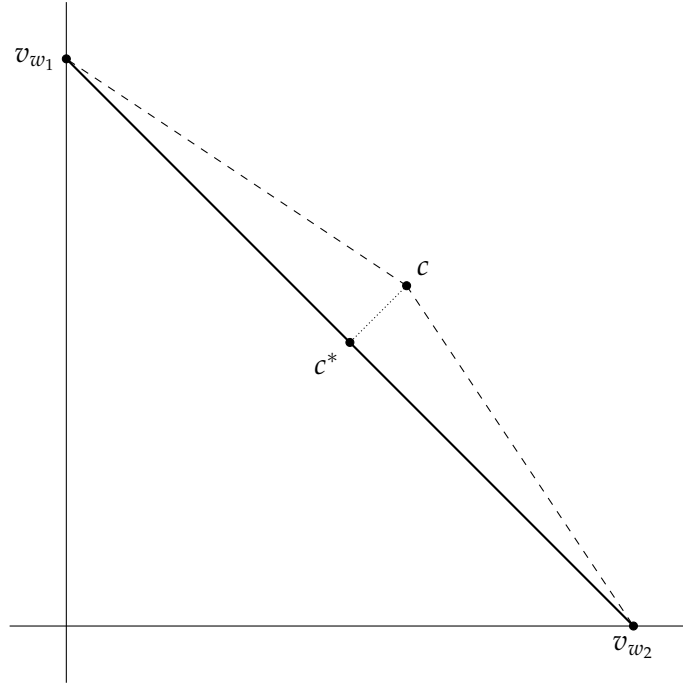


Figure 1: If a credence function is defined only on propositions  $A$  and  $B$ , we can represent it by the point on the Cartesian plane that takes the credence assigned to  $A$  as its  $x$ -coordinate and the credence assigned to  $B$  as its  $y$ -coordinate. Thus, in the diagram, credence function  $c$  assigns 0.6 to both  $A$  and  $B$ , whereas credence function  $c^*$  assigns 0.5 to both. The two ideal credence functions in this situation are  $v_{w_1}$  and  $v_{w_2}$ : the former is ideal at worlds at which  $A$  is false and  $B$  is true; the latter at worlds at which  $A$  is true and  $B$  is false. The credence functions that satisfy (Prob<sub>2</sub>) are the ones that lie on the line that joins  $v_{w_1}$  and  $v_{w_2}$ . Thus, since  $c$  lies off this line, it violates (Prob<sub>2</sub>), while  $c^*$  satisfies it. Having represented credence functions in this way, a credence function has a greater Brier score at a world the further the point that represents it lies from the point that represents that world's ideal credence function. It is clear by Pythagoras' Theorem that  $c^*$  lies closer to  $v_{w_1}$  than  $c$  does; and  $c^*$  lies closer to  $v_{w_2}$  than  $c$  does. That is,  $c^*$  strongly Brier dominates  $c$ , as Theorem 1 demands.

Therefore,

(4) (Prob<sub>2</sub>)

I wish to offer a similar style of argument in favour of the Principle of Indifference. Here, I will state that argument in a restricted version that attempts only to establish a particular instance of the Principle of Indifference; it is the instance that applies to the sort of agent we've been considering in this section, namely, an agent with opinions about only the propositions  $A$  and  $B$ . Before we state the norm, we define a particular credence function: let  $c_0(A) = c_0(B) = 0.5$ .

(PoI<sub>2</sub>) At the beginning of an agent's credal life, if she entertains only  $A$  and  $B$ , it ought to be the case that her credence function is  $c_0$ .

Note that this norm applies only to Lewis' superbabies. This distinguishes it from (Prob<sub>2</sub>), which applies to an agent at any time in her credal life.

I will retain Joyce's first premises, (1)(i) and (1)(ii)—that is, his account of cognitive value. But I will replace Dominance with an alternative norm of decision theory, namely, the Minimax rule, which demands that an agent at the beginning of her cognitive life choose one of the options that minimizes its maximum (or worst-case) disutility.<sup>4</sup> More precisely:

(Minimax) Suppose  $\mathcal{O}$  is a set of options,  $\mathcal{W}$  is a set of possible worlds, and  $U$  is a measure of the value of the options in  $\mathcal{O}$  at the worlds in  $\mathcal{W}$ . Suppose  $o, o'$  in  $\mathcal{O}$ . Then we say

(c)  $o$  *worst-case dominates*  $o'$  *with respect to*  $U$  if the minimum utility of  $o$  is greater than the minimum utility of  $o'$ . That is,

$$\min_{w \in \mathcal{W}} U(o, w) > \min_{w \in \mathcal{W}} U(o', w)$$

Now suppose  $o, o'$  in  $\mathcal{O}$  and

- (i)  $o$  worst-case dominates  $o'$
- (ii) There is no  $o''$  in  $\mathcal{O}$  such that  $o''$  worst-case dominates  $o$

Then, for an agent at the beginning of her cognitive life,  $o'$  is irrational.

As above, note that the norm applies only to Lewis' superbabies. This distinguishes it from the norm (Dominance) to which Joyce's argument appeals; (Dominance) applies to an agent at any time in her credal life.

---

<sup>4</sup>This norm is sometimes known as the Maximin rule, under which guise it demands (equivalently) that an agent choose one of the options that maximises its minimum (or worst-case) utility.

Again, we ask: which credal states does (Minimax) rule out? This is answered by the following theorem:

**Theorem 2** *Define  $c_0$  as follows:  $c_0(A) = c_0(B) = 0.5$ . Then, for all credence functions  $c \neq c_0$  defined for  $A$  and  $B$  only,*

$$\max_{w \in \mathcal{W}} B(c_0, w) < \max_{w \in \mathcal{W}} B(c, w)$$

*Equivalently,*

$$\min_{w \in \mathcal{W}} -B(c, w) < \min_{w \in \mathcal{W}} -B(c_0, w)$$

With this in hand, we can state the following argument for the particular case of the Principle of Indifference we have been considering:

- (1) The cognitive value of a credence function  $c$  and  $w$  is given by  $-B(c, w)$ , the negative of the Brier score of  $c$  at  $w$ .
- (2) (Minimax)
- (3) Theorem 2

Therefore,

- (4) (PoI<sub>2</sub>)

Theorem 2 is a corollary of Theorem 4, which we state and prove below. Figure 2 illustrates the theorem.

Thus, according to the Argument from Accuracy, what is wrong with an agent who violates (PoI<sub>2</sub>) is that she risks greater inaccuracy than she needs to risk. If she adopts  $c_0$ , her inaccuracy in the worst-case scenario (as measured by the Brier score) is 0.25. If instead she adopts  $c(A) = 0.2$  and  $c(B) = 0.8$ , for instance, her worst-case inaccuracy is 1.28. And similarly for any other  $c \neq c_0$  (though with different numerical values): her worst-case scenario if she opts for  $c$  is worse than the worst-case scenario of an agent who opts for  $c_0$ .

In section 5.1, we will generalise this argument by showing that it applies to agents with credences over larger sets of propositions than  $\{A, B\}$ ; and in section 5.2, we will strengthen it by showing that the Principle of Indifference still follows if we replace the Brier score with any of a vast array of alternative measures of inaccuracy.

### 3 Existing justifications of the Principle of Indifference

For the moment, I will continue to focus on this simple version of the Argument from Accuracy in order to assess it. First, I will show that it overcomes



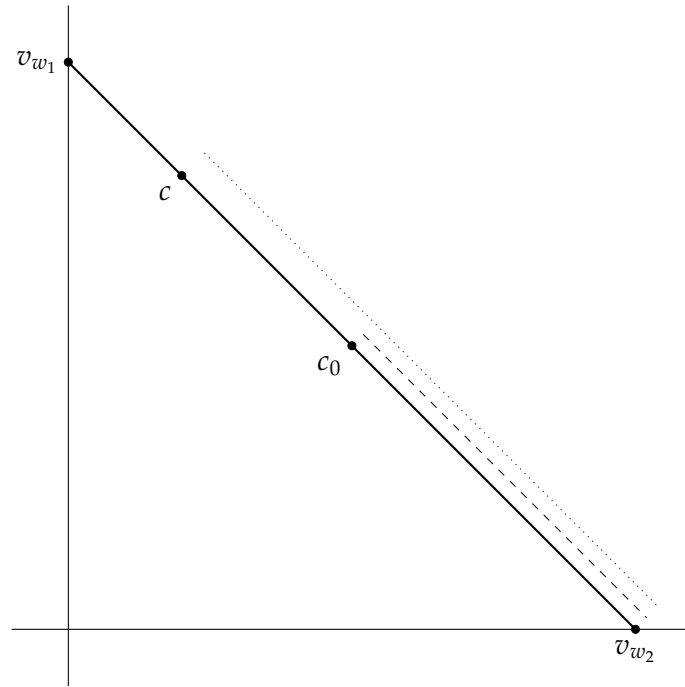


Figure 2: We represent credence functions on the Cartesian plane as above. As above,  $c_0(A) = c_0(B) = 0.5$ . And  $c(A) = 0.2$ ;  $c(B) = 0.8$ . The dashed line represents the maximum distance of  $c_0$  from the omniscient credence function, and thus the maximum disutility that  $c_0$  risks—it achieves that maximum at both worlds. The dotted line, on the other hand, represents the maximum distance of  $c$  from the omniscient credence function, and thus the maximum disutility that  $c$  risks—it achieves that maximum when  $A$  is true and  $B$  is false. Clearly, the maximum disutility risked by  $c$  is greater than the maximum disutility risked by  $c_0$ .

the problems faced by the three main attempts to justify the Principle of Indifference. Next, I will consider further problems that it may face. Thus, throughout, we are considering a superbaby with opinions about only the exclusive and exhaustive pair of propositions  $A$  and  $B$ .

### 3.1 The Argument from Evidential Support

The argument (cf. [White, 2009, §4.3]):

An agent ought to believe a proposition exactly to the extent that her evidence supports it. A superbaby has no evidence. An empty body of evidence supports a pair of exclusive and exhaustive propositions equally. Therefore, our agent ought to have equal credence in  $A$  and in  $B$ . In the presence of (Prob<sub>2</sub>), this gives that our agent ought to adopt  $c_0$ . Therefore, (PoI<sub>2</sub>).

First problem: Our best account of comparative evidential support—namely, Bayesian confirmation theory—is given in terms of the effect of evidence on a rational agent’s credences. So, in order to know that a body of evidence supports two hypotheses equally, and thus to mobilise this argument, we already need to know all the principles that govern a rational agent’s credences. Thus, we cannot establish the first premise of the Argument from Evidential Support without already knowing whether the Principle of Indifference is true or not. So the argument cannot be used to adjudicate upon the truth of this principle.

Second problem: It is not clear *why* an agent should match her credence to the evidential support in the way the first premise of this argument demands. What good does she gain by doing this? Of course, one might hold that responding correctly to the evidence is an irreducible, fundamental cognitive good. The problem with this is that it cannot be the only such good. As Alvin Goldman points out, if responding correctly to one’s evidence were the only cognitive virtue, we would have no cognitive reason to gather new evidence, since we could knowingly obtain maximal cognitive value while in possession of a very impoverished evidential base [Goldman, 2002, §3]; and I would do my students no cognitive harm by making them believe false propositions about Kant’s life on the basis of my testimony, given that they are justified in trusting me on such matters [Goldman, 2002, §6]. Thus, even if responding correctly to one’s evidence is a fundamental cognitive good, accuracy must be one too. But we then have two fundamental cognitive goods that might come into conflict. Thus, we will need an account of how they ought to be weighed against one another. Of course, it is not impossible to give such an account, but it is a burden that such a justification has to bear, while my alternative does not. Indeed, one of the appealing features of the Argument from Accuracy is that it is compatible with *cognitive value monism*, the view that there is only

one source of cognitive value; it is compatible with the version of cognitive value monism that Goldman dubs *veritism*, namely, that accuracy is the only source of cognitive value. Cognitive value monism does not have to say how competing cognitive values ought to be weighed, since it recognises only one source of such value.

Now, it might be objected that, just as the good of having accurate credences cannot be reduced to the good of matching those credences to the evidential support, so the good of matching credences to evidential support cannot be reduced to the good of accuracy. That is, cognitive value monism must be wrong. I see the Argument from Accuracy presented in this paper as a part of a larger project to answer this worry. The aim is precisely to show that we can establish important evidential norms by appealing only to the good of accuracy, and thereby reduce the virtue of responding appropriately to the evidence to the virtue of accuracy. The Principle of Indifference is one such norm—it says how an agent should respond in the absence of evidence. I take this to be the project begun in [Joyce, 1998] and continued by [Easwaran, 2013] and [Pettigrew, 2013]. One of the great advantages of this view is its potential to unify explanation in epistemology: all cognitive norms are justified and explained by appeal to a single cognitive virtue, namely, accuracy.

### 3.2 The Argument from Minimal Information

The argument:

An agent's credences ought to incorporate the information given by her evidence, but no more. Thus, an agent with no evidence ought to have credences that contain minimal information. These are given by  $c_0$ . Therefore, (PoI<sub>2</sub>).

There are two ways to make this argument:

- (1) Assume as a normative principle that credences should incorporate the information given by evidence, but no more. Then provide a precise measure of the information contained in a set of credences and show that the credences that incorporate minimal information are those demanded by the Principle of Indifference.

This is Jaynes' approach [Jaynes, 1957]. He appeals to Shannon's measure of the information contained in a probability distribution: the entropy of a distribution measures its lack of informational content.

[I]n making inferences on the basis of partial information we must use that probability distribution which has maximum entropy subject to whatever is known. This is the only unbiased assignment we can make; to use any other

would amount to arbitrary assumption of information which by hypothesis we do not have. [...] The maximum entropy distribution may be asserted for the positive reason that it is uniquely determined as the one that is maximally noncommittal with regard to missing information. [Jaynes, 1957, 623]

- (2) Lay down conditions on a rule  $N$  that assigns to each body of evidence  $S$  a unique set of credences  $N(S)$  that it is rational to have in the presence of that evidence. Incorporate into these conditions one that says that  $N(S)$  should not contain information beyond that contained in  $S$ . Then show that only one rule satisfies those conditions and that, when  $S$  is the empty body of evidence,  $N(S)$  is the set of credences demanded by the Principle of Indifference.

This is the approach taken by Paris and Vencovská [Paris and Vencovská, 1990]. They list amongst their conditions on  $N$  one that they gloss as

“ $N(S)$  should not make any assumptions beyond those contained in  $S$ .” [Paris and Vencovská, 1990, 185]

In both cases, the problem is the same: Neither argument tells us why it is bad to have credences that incorporate more information than is demanded by one’s evidence. Other things being equal, it is presumably better to have more informative credences. Thus, if we are to minimise the informativeness of our credences as much as possible without violating our evidential obligations, it must be because by adopting more informative credences we risk something bad. What the Argument from Minimal Information lacks in all its incarnations is an account of the badness we risk.

Of course, the Argument from Accuracy addresses this concern. What one risks by adopting credences other than those mandated by  $(PoI_2)$  is greater inaccuracy. So, while having greater informational content is clearly valuable (though not a fundamental good), if one has credences that encode more informational content than one’s evidence demands, there will be a world at which one’s inaccuracy is greater than the inaccuracy at any world of the credences mandated by  $(PoI_2)$ . Thus, any gain in information content comes with a risk attached. This is why we should minimise informational content.

In sum: in order to save the Argument from Minimal Information, we had to appeal to the claim that we ought not to risk greater-than-necessary inaccuracy. But given that it is this claim that grounds the Argument from Accuracy, surely it is better to dispense with the talk of informational content altogether and appeal to accuracy directly in our justification of  $(PoI_2)$ .

### 3.3 The Argument from Caution

The argument:

In the face of risky practical decisions between different actions, an agent ought to be cautious. On average, in the face of risky practical decisions, the credences given by  $c_0$  lead an agent to make the most cautious choices. Therefore, (PoI<sub>2</sub>).

Jon Williamson proposes this argument, making precise what he means by a risky decision and giving a measure of how cautious a particular set of credences is [Williamson, 2007, §8].

One might object to this argument in the way that many object to the Dutch Book argument for Probabilism: it is too limited in its scope. Probabilism, if true, is true not only for an agent who knows she will meet a Dutch Bookie. If true, it is true also for agents who do not know what sort of decisions they will face, or who know that they will not encounter a Dutch Bookie. But, the objection goes, the Dutch Book argument has no power against such an agent. Similarly, the Principle of Indifference, if true, is true for any agent, regardless of what she knows about the practical decisions she'll face; regardless of what she knows about how risky those decisions will be. But, the objection goes, Williamson's Argument from Caution has no power against an agent who knows she will never face a risky decision.

However, just as we can 'depragmatize' the Dutch Book argument for Probabilism, so we can 'depragmatize' the Argument from Caution.<sup>5</sup> The depragmatized Dutch Book argument says that an agent who violates Probabilism is irrational because there are choices that lead to sure losses that her credences will *sanction*, whether or not she will ever be in a position to make those choices [Christensen, 1996]. Similarly, a depragmatized Argument from Caution says that an agent who violates the Principle of Indifference is irrational because there are incautious choices in response to risky decisions that her credences *sanction*, whether or not she will ever face those decisions.

My concern with Dutch Book arguments and the Argument from Caution—whether pragmatic or depragmatized—is that they identify only part of what is wrong with an agent who violates the norm they seek to establish. They establish that violating the norm has bad consequences for an agent's *practical* reasoning—she will make bad choices or she will be committed to making bad choices between different actions. But violating these norms seems to involve a failure of *theoretical* reason.<sup>6</sup> Thus, if possible, we would

---

<sup>5</sup>Thanks to an anonymous referee for pushing me to consider this move.

<sup>6</sup>Having said that, [Skyrms, 1984, 21-26] and [Armendt, 1993] both present Dutch Book arguments as dramatising failures of theoretical reason—they take themselves to be following Ramsey in this respect. But I share Vineberg's concerns about these interpretations

like to justify them in a way that appeals only to theoretical reason and its goals. And, indeed, that is what the Argument from Evidential Support and the Argument from Minimal Information seek to do. But I have argued that both of these fail. Now, it may be simply impossible to provide such a justification; there may be no purely cognitive justification available. In that case, we would have discovered something interesting: these norms, which seem to govern theoretical reason in the first place and practical reason only secondarily, in fact govern practical reason in the first place and theoretical reason only secondarily. But I claim that it is possible; and I will argue for that by presenting the Argument from Accuracy, which attempts to provide a purely cognitive justification for the Principle of Indifference. Thus, I do not offer an objection to the Argument from Caution, but rather a motivation for looking beyond to it to find a purely cognitive justification for the Principle of Indifference.

## 4 Minimax and risk aversion

These, then, are the merits of the Argument from Accuracy: it appeals only to the cognitive value of accuracy and is thus compatible with the parsimonious and explanatorily powerful thesis of cognitive value monism; and it is entirely non-pragmatic, so it reveals the purely cognitive reason for obeying the Principle of Indifference. Its primary demerit is that it relies on Minimax, which many will say is not a norm of rational choice. What, then, can be said in favour of this assumption?

Let me begin by noting that, as I have stated it, Minimax has a very limited domain of application: it applies only at the beginning of an agent's credal life, before she has acquired any evidence and before she has assigned credences to the propositions she entertains; indeed, according to the Argument from Accuracy, it is the norm that dictates how an agent without credences ought to assign her initial credences. For an agent at any other stage of her credal life, Minimax does not apply. Instead, in those situations, the agent ought to maximise her subjective expected utility relative to the credences she has at that time (or perhaps her risk-weighted expected utility [Buchak, 2013]); and this will often lead to choices that Minimax would condemn as irrational (at least providing the risk weighting is not extreme).

Thus, the version of Minimax to which I appeal is akin to the version deployed by Rawls in *A Theory of Justice* [Rawls, 1975, Chapter III]. Agents at the beginning of their life are in a credal version of the 'original position'; they behind a credal version of Rawls's 'veil of ignorance'. Just as a Rawlsian agent must choose the society in which she will live in accordance

---

[Vineberg, 2001].

with Minimax, I claim that a Lewisian superbaby ought to adopt her initial credence function in accordance with that decision-theoretic principle.

However, while I may allay some fears about Minimax by noting the weakness of the version to which I appeal, others remain. After all, there are other putative decision-theoretic norms that purport to govern agents who have yet to assign credences. The norm of (Dominance) to which Joyce appeals is such a norm. And there are other putative norms of this sort that are supposed to capture the rationality of risk-sensitive behaviour: Minimax makes all but the most risk-averse behaviour irrational; Maximax makes all but the most risk-seeking behaviour irrational. Minimax demands that an agent ought to adopt credence function  $c_0$ . Maximax demands that she ought to adopt  $v_{w_1}$  or  $v_{w_2}$ , both of which achieve maximal accuracy; the former at  $w_1$ , the latter at  $w_2$ . Why favour Minimax over Maximax?

The answer, I think, lies in cognitive conservatism. As William James noted,

“There are two ways of looking at our duty in the matter of opinion,—ways entirely different, and yet ways about whose difference the theory of knowledge seems hitherto to have shown very little concern. We must know the truth; and we must avoid error,—these are our first and great commandments as would-be knowers.” [James, 1905, §VII]

But, as he also noted, these two laws are in tension. Which has the upper hand in disputes? For the cognitive conservative, it is the latter; for the cognitive radical, it is the former. I side with the conservative.

I have no argument for making this alliance. At this point, it seems to me, we have reached normative bedrock: one cannot argue for cognitive conservatism from more basic principles. Thus, to those who reject cognitive conservatism—and it seems that James himself wished to do so—I can only recommend the argument of this paper as an argument for the following (subjunctive) conditional: Cognitive Conservatism  $\Rightarrow$  Principle of Indifference.

In the accuracy framework that is the setting for Joyce’s argument and mine, James’ “duty” to “know the truth” becomes the goal of having highly accurate credences, and the “duty” to “avoid error” becomes the goal of not having highly inaccurate credences. The tension to which James alludes arises because, in a state of ignorance, one pursues the former goal by having credences close to the omniscient credences at some possible world, which are therefore highly accurate at that world; however, such credences will be very far away from the omniscient credences at any other possible world, and thus highly inaccurate there. So one can fully pursue one goal only by ignoring the other. Our decision is thus whether to leave open the possibility of “knowing the truth” (that is, being highly accurate)

and thereby incurring a risk of being in “error” (that is, being highly inaccurate); or ensure that we “avoid error” by having credences that are not highly inaccurate at any world, but thereby preclude the possibility of “knowing the truth”, since such credences will also not be highly accurate at any world. Opting for the former—“knowing the truth”—amounts to following a risk-seeking norm; opting for the latter—“avoid error”—amounts to following a risk-averse norm.

Now, it follows from this that the cognitive conservative ought to adopt a risk-averse norm. But it does not tell between the different risk-averse norms that have been formulated. I will not attempt to consider all such norms here. I only wish to note that, on one of the most plausible rivals to Minimax—namely, Minimax Regret—the Argument from Accuracy still goes through. I leave consideration of more sophisticated risk-averse norms, such as various versions of the Hurwicz criterion, to another time.

The intuition behind Minimax Regret is this: What you ought to minimise is not the maximum *disutility* that you risk by choosing a particular option, as Maximin demands; rather you ought to minimise the maximum *regret* you risk, where the regret that attaches to an option at a world is the difference between the maximum utility that attaches to some option at that world and the utility of the option in question.<sup>7</sup> Thus:

(Minimax Regret) Suppose  $\mathcal{O}$  is a set of options,  $\mathcal{W}$  is a set of possible worlds, and  $U$  is a measure of the value of the options in  $\mathcal{O}$  at the worlds in  $\mathcal{W}$ . Given an option  $o$  in  $\mathcal{O}$ , define the *regret generated by choosing  $o$  at  $w$*  to be the difference between the maximum utility that could have been obtained by one of the options at  $w$  and the utility of  $o$  at  $w$ :

$$R_U(o, w) := \max_{o' \in \mathcal{O}} U(o', w) - U(o, w)$$

Suppose  $o, o'$  in  $\mathcal{O}$ . Then we say

- (c)  $o$  *worst-case regret-dominates*  $o'$  *with respect to*  $U$  if the maximum regret generated by choosing  $o$  is less than the maximum regret generated by picking  $o'$ . That is,

$$\max_{w \in \mathcal{W}} R_U(o, w) < \max_{w \in \mathcal{W}} R_U(o', w)$$

Now suppose  $o, o'$  in  $\mathcal{O}$  and

- (i)  $o$  worst-case regret-dominates  $o'$
- (ii) There is no  $o''$  in  $\mathcal{O}$  such that  $o''$  worst-case regret-dominates  $o$

---

<sup>7</sup>This decision-theoretic principle was first considered in [Savage, 1951].



Then  $o'$  is irrational.

Now, in many cases, the *regret* that attaches to an option  $o$  at a world  $w$  is different from the *disutility* of  $o$  at  $w$ ; and often that difference leads (Minimax) and (Minimax Regret) to rule out different options as irrational. However, in the case of credence functions, the maximum cognitive utility at a world  $w$  is 0, for any  $w$ —this maximum is obtained uniquely by  $v_w$ . Thus, the regret that attaches to a credence function  $c$  at  $w$  is the difference between the maximum cognitive utility at  $w$ —which is 0, as we have just seen—and  $-B(c, w)$ . But this difference is, of course,  $B(c, w)$ , which is the cognitive disutility of  $c$  at  $w$ . Thus, the regret  $R_{-B}(c, w)$  that attaches to  $c$  at  $w$  is just the disutility  $B(c, w)$  of  $c$  at  $w$ . As a result, when the options in question are credence functions and the utility function is an accuracy measure, (Minimax) and (Minimax Regret) rule out exactly the same options as irrational. Thus, since (Minimax) entails (PoI<sub>2</sub>), so does (Minimax Regret).

## 5 Extending the Argument from Accuracy

So far, we have justified only a particular instance of the Principle of Indifference, which says that an agent with no evidence and who has credence in only two exclusive and exhaustive propositions ought to assign equal credence to each of them. What is the general principle from which this follows? In this section, we formulate a general version of the Principle of Indifference, present the Argument from Accuracy in its favour, and then see how we may strengthen it by weakening the first premise.

### 5.1 Generalizing the argument: expanding the algebra

As before, we seek a norm that governs superbabies. Such agents have acquired no evidence and they wish to assign credences to the propositions they entertain. Let  $\mathcal{F}$  be the set of propositions that our agent entertains. Thus, hitherto, we have assumed that  $\mathcal{F}$  is the two-element set  $\{A, B\}$ . Henceforth, we will assume that  $\mathcal{F}$  forms a finite algebra of propositions—thus,  $\mathcal{F}$  is closed under all Boolean operations.<sup>8</sup> Now let us say that the *possible worlds relative to  $\mathcal{F}$*  are the classically consistent assignments of truth values to the elements of  $\mathcal{F}$ —denote the set of these worlds  $\mathcal{W}_{\mathcal{F}}$ . These are the different possibilities grained as finely as the propositions to which the agent assigns credences will allow. Thus, in our example,  $\mathcal{W}_{\mathcal{F}} = \{w_1, w_2\}$  where

---

<sup>8</sup>Of course,  $\{A, B\}$  is not itself an algebra since it misses out the top and bottom elements  $\top$  and  $\perp$ . But Dominance already demands unique values for  $\top$  (namely, 1) and  $\perp$  (namely, 0). So Minimax does as well. Thus, given that Minimax demands  $c(A) = c(B) = 0.5$  for an agent with credences only over  $\{A, B\}$ , we can infer that it demands  $c(\top) = 1$ ,  $c(A) = c(B) = 0.5$ , and  $c(\perp) = 0$  for an agent with credences over  $\{\top, A, B, \perp\}$ .

- $w_1(A) = \text{true}, w_1(B) = \text{false};$
- $w_2(A) = \text{false}, w_2(B) = \text{true}.$

Since  $\mathcal{F}$  is finite, so is  $\mathcal{W}_{\mathcal{F}}$ .

With this in hand, we can state the version of the Principle of Indifference I wish to justify: in the absence of any evidence, an agent's credence in a proposition ought to be the proportion of the possible worlds relative to  $\mathcal{F}$  at which that proposition is true. Thus, define the credence function  $c_0^{\mathcal{F}}$  as follows: for all  $X$  in  $\mathcal{F}$

$$c_0^{\mathcal{F}}(X) := \frac{|\{w \text{ in } \mathcal{W}_{\mathcal{F}} : w(X) = \text{true}\}|}{|\{w \text{ in } \mathcal{W}_{\mathcal{F}}\}|}$$

Then we state the Principle of Indifference as follows:

(PoI) At the beginning of an agent's credal life, if she entertains only the propositions in  $\mathcal{F}$ , it ought to be the case that her credence function is  $c_0^{\mathcal{F}}$ .

And here is the Argument from Accuracy that I propose in its favour:

- (1) The cognitive value of a credence function is given by the negative of its Brier score at that world.
- (2) (Minimax)
- (3) **Theorem 3** For all credence functions  $c$  on  $\mathcal{F}$ , if  $c \neq c_0^{\mathcal{F}}$ , then

$$\max_{w \text{ in } \mathcal{W}_{\mathcal{F}}} B(c_0^{\mathcal{F}}, w) < \max_{w \text{ in } \mathcal{W}_{\mathcal{F}}} B(c, w)$$

Therefore,

- (4) (PoI)

Theorem 3 follows from Theorem 4, which we prove below.

## 5.2 Strengthening the argument: beyond the Brier score

The argument of the preceding section generalizes our earlier justification to cover a much wider range of agents; now we strengthen that argument. We can weaken premise (1) considerably and nonetheless retain our conclusion (PoI). As we will see, Theorem 3 still holds if we replace the Brier score with any of a vast array of alternative measures of inaccuracy. Let's see what they are.

In this section, we fix  $\mathcal{F}$  and drop the superscript on ' $c_0^{\mathcal{F}}$ ', writing ' $c_0$ ' in its place. In section 7, we will return to the fact that the credence function demanded by the Principle of Indifference is sensitive to the set  $\mathcal{F}$  of propositions that the agent entertains.

### 5.2.1 Conditions on inaccuracy measures

What features do we require of a measure of inaccuracy? I will require two. First, I suggest that the inaccuracy of a credence function at a world ought to be blind to the content of the propositions for which that credence function is defined. Thus, if I have credence only in *3 is prime* and my brother has credence only in *7 is prime* and we each have credence 0.9 in our respective proposition, then an inaccuracy measure ought to render us equally accurate, since both propositions are true. Similarly, if we both have credences in both propositions, but I have credence 0.9 in the former and 0.8 in the latter and my brother has 0.8 in the former and 0.9 in the latter, then an inaccuracy measure ought to render us equally accurate, since both are true. Thus, the inaccuracy of a credence function at a world ought to depend on the pairs  $(c(X), v_w(X))$  of its credence in  $X$  together with the omniscient credence in  $X$ , but not on the content of the propositions that give rise to those pairs. I will, however, allow that an inaccuracy measure can treat propositions of different strengths differently. That is, while an inaccuracy measure must be blind to content, it need not be blind to logical or modal strength. Let's make all of this precise:

Suppose:

- $c, d$  are credence functions defined on finite algebras  $\mathcal{F}, \mathcal{G}$  respectively;
- $w$  is a world in  $\mathcal{W}_{\mathcal{F}}$ ; and  $u$  is a world in  $\mathcal{W}_{\mathcal{G}}$ .

Now suppose that there is an isomorphism  $\tau : \mathcal{F} \cong \mathcal{G}$  such that, for all  $X \in \mathcal{F}$ ,

$$c(X) = d(\tau(X)) \quad \text{and} \quad v_w(X) = v_u(\tau(X))$$

Then it ought to be the case that  $c$  is exactly as inaccurate at  $w$  as  $d$  is at  $u$ —that is,  $I(c, w) = I(d, u)$ . We say that a function with this property is *egalitarian*.<sup>9</sup>

Second, I suggest that an inaccuracy measure should make the following true: there is no credence function  $c \neq c_0$  that is at least as accurate as  $c_0$  at all possible worlds. We say that a function with this property *renders indifference immodest*. To justify this claim, we adapt an argument that Jim Joyce has advanced in favour of a related requirement on inaccuracy measures [Joyce, 2009, §9]. Our first pass at this argument is based on four claims:

---

<sup>9</sup>The Brier score is egalitarian. So are all the inaccuracy measures generated by a certain sort of scoring rule called a *proper scoring rule*. For more on proper scoring rules and the sort of inaccuracy measures they generate, see [Predd et al., 2009].

- (1) There is a possible world at which  $c_0$  gives the chances of the propositions in  $\mathcal{F}$  (this is a weak version of Joyce's claim).
- (2) If an agent knows the chances and nothing more, her credences should match the chances (this is a version of Lewis' Principal Principle).
- (3) Accuracy is the sole cognitive virtue (this is cognitive value monism, as discussed in section 2)
- (4) If an option  $o$  is rationally permissible, and another option  $o'$  has at least as great utility as  $o$  at every world, then option  $o'$  is rationally permissible as well (this is a weak version of Dominance).

The argument then runs as follows:

By (1), there is a possible world at which  $c_0$  gives the objective chances of the propositions in  $\mathcal{F}$ . Thus, by (2), there is an evidential situation such that  $c_0$  is the unique rational response to that situation, namely, the situation in which an agent learns that the objective chances are given by  $c_0$  (and learns nothing more). Now, suppose it is the case that there is another credence function  $c$  such that  $c$  is at least as accurate as  $c_0$  at all worlds. Then, since accuracy is the single fundamental virtue of credence functions (3),  $c$  has at least as great cognitive utility as  $c_0$  at every world. So, by (4), it would be rationally permissible to adopt  $c$  rather than  $c_0$  in the situation in question. But, by assumption, this is not the case— $c_0$  is the unique rational response to the evidence. Therefore, there is no such credence function  $c$ , as required.

Of course, one might wonder why a weak version of Dominance is appropriate as premise (4). Should we not appeal to a version of Minimax instead? But note that the evidential situation in which  $c_0$  is the unique rational response is not the evidential situation of a superbaby—it is a situation in which the agent has acquired considerable evidence. And Minimax, we have noted above, applies only to superbabies.

Alan Hájek has criticised assumption (1) as it appears in Joyce's original argument [Hájek, 2008, 814]. He objects that, depending on the content of the propositions in  $\mathcal{F}$ , it may be that  $c_0$  could not be the chance function of any world. For instance,  $\mathcal{F}$  may contain propositions to which chances cannot be ascribed: moral propositions, aesthetic propositions, mathematical propositions, or propositions about the chances themselves. Any of these may be appropriate objects of belief, but not propositions to which chances might be ascribed.

However, we can answer this objection as follows. We replace (1) by:

(1a) For any finite algebra  $\mathcal{F}$ , there is a finite algebra  $\mathcal{G}$  and an isomorphism  $\tau : \mathcal{F} \cong \mathcal{G}$  such that, if we define  $d_0(\cdot) := c_0(\tau^{-1}(\cdot))$ , then  $d_0$  gives the chances of the propositions in  $\mathcal{G}$  at some world.

(1b) Inaccuracy measures ought to be egalitarian.

(1a) says that, although the content of the propositions in  $\mathcal{F}$  might prevent  $c_0$  on  $\mathcal{F}$  from being a chance function, there is always an isomorphic algebra  $\mathcal{G}$  such that the isomorphic copy  $d_0$  of  $c_0$  on  $\mathcal{G}$  could be a chance function. Thus, by (2), there is an evidential situation that demands  $d_0$ . And thus there must be no  $d$  defined on  $\mathcal{G}$  such that  $I(d, u) \leq I(d_0, u)$  for all worlds  $u \in \mathcal{W}_{\mathcal{G}}$ . But, by (1b), since  $I$  is egalitarian, we have that there is no  $c$  defined on  $\mathcal{F}$  such that  $I(c, w) \leq I(c_0, w)$  for all worlds  $w \in \mathcal{W}_{\mathcal{F}}$ , as required.

So how do we justify (1a)? Suppose  $\mathcal{F}$  is an algebra and  $|\mathcal{W}_{\mathcal{F}}| = n$ . Then consider the following  $n$  exclusive and exhaustive propositions: *This  $n$ -sided die will land on 1, This  $n$ -sided die will land on 2, ..., This  $n$ -sided die will land on  $n$* . Let  $\mathcal{G}$  be the algebra generated by these propositions. And let  $\tau$  be an isomorphism between  $\mathcal{F}$  and  $\mathcal{G}$ . Then let  $d_0(\cdot) := c_0(\tau^{-1}(\cdot))$ . How do we know that  $d_0$  is the chance function at some world? We know there is a possible world at which the  $n$ -sided die in question is fair. At that world,  $d_0$  gives the chances. Thus, from (1a), (1b), (2), and (3), it follows that our inaccuracy measure ought to render indifference immodest.<sup>10</sup>

However, there is another concern with this argument: it arises because we appealed to the Principal Principle in (2) and cognitive value monism in (3). But the Principal Principle seems to be an evidentialist norm, and one might worry that cognitive value monism cannot justify it. However, as [Pettigrew, 2013] shows, this is not the case: there is an accuracy-based argument for the Principal Principle that has the same structure as Joyce's argument for Probabilism and the Argument from Accuracy for the Principle of Indifference that I am defending here. So it is not at odds with the version of cognitive value monism defended here, namely, veritism.

### 5.2.2 The Argument from Accuracy strengthened

Thus, we require that an inaccuracy measure should be egalitarian and that it should render indifference immodest. Theorem 4 below shows that, for any such inaccuracy measure,  $c_0$  is the only credence function that minimises maximum inaccuracy. Roughly, the reason is this: First, since  $I$  is egalitarian and  $\mathcal{F}$  is an algebra,  $c_0$  is equally inaccurate at all possible worlds relative to  $\mathcal{F}$ . Thus,  $c_0$  obtains its maximum inaccuracy at every world. Second, since  $I$  renders indifference immodest, if  $c \neq c_0$ , then there

<sup>10</sup> Again, the Brier score renders indifference immodest, as does any inaccuracy measure generated by a proper scoring rule. For a proof, see [Predd et al., 2009, §6].

is some world at which  $c$  is more inaccurate than  $c_0$ . Since  $c_0$  obtains its maximum inaccuracy at that world (as it does at every world), it follows that the maximum inaccuracy of  $c$  is greater than that of  $c_0$ . So (Minimax) rules out  $c$  as irrational. More formally:

**Theorem 4** *Suppose  $I$  is egalitarian and renders indifference immodest. Then, for all credence functions  $c$  on  $\mathcal{F}$ , if  $c \neq c_0$ , then*

$$\max_{w \in \mathcal{W}_{\mathcal{F}}} I(c_0, w) < \max_{w \in \mathcal{W}_{\mathcal{F}}} I(c, w)$$

*Proof.* We begin by proving that, if  $I$  is egalitarian, then, for any  $w, w' \in \mathcal{W}_{\mathcal{F}}$ ,

$$I(c_0, w) = I(c_0, w')$$

We will show that, for any  $w, w' \in \mathcal{W}_{\mathcal{F}}$ , there is an automorphism  $\tau : \mathcal{F} \cong \mathcal{F}$  such that

$$c_0(X) = c_0(\tau(X)) \quad \text{and} \quad v_w(X) = v_{w'}(\tau(X))$$

Define  $\tau$  as follows: Represent propositions as sets of possible worlds. Then define the following permutation on the set of possible worlds  $\mathcal{W}_{\mathcal{F}}$ :  $\tau$  swaps  $w$  and  $w'$  and leaves everything else fixed. Then, finally, define the following automorphism on  $\mathcal{F}$ :

$$\tau(X) := \{\tau(w) : w \in X\}.$$

Then  $\tau$  has the required properties, since  $|\{w \in \mathcal{W}_{\mathcal{F}} : w(X) = \text{true}\}| = |\{w \in \mathcal{W}_{\mathcal{F}} : w(\tau(X)) = \text{true}\}|$ . So, since  $I$  is egalitarian, we have  $I(c_0, w) = I(c_0, w')$ .

Now, since  $I$  renders indifference immodest, it follows that, for every credence function  $c$ , there is a world  $w \in \mathcal{W}_{\mathcal{F}}$  such that  $c$  is more inaccurate than  $c_0$  at  $w$ : that is,  $I(c_0, w) < I(c, w)$ . But, by the previous result, it follows that, for any world  $w' \in \mathcal{W}_{\mathcal{F}}$ ,  $c$  is more inaccurate at  $w$  than  $c_0$  is at  $w'$ : that is,  $I(c_0, w') = I(c_0, w) < I(c, w)$ . So it follows that

$$\max_{w \in \mathcal{W}_{\mathcal{F}}} I(c_0, w) < \max_{w \in \mathcal{W}_{\mathcal{F}}} I(c, w)$$

as required. □

This gives us an extremely general version of the Argument from Accuracy. It establishes the Principle of Indifference for agents with credences in propositions that form a finite algebra. And it assumes only two plausible properties of inaccuracy measures, as well as the decision-theoretic norm of Minimax.

## 6 Other norms

Having completed our argument for the Principle of Indifference, we ask how this norm interacts with other Bayesian norms that are taken to govern superbabies. These are: Probabilism, Regularity, and the Principal Principle.

### 6.1 (PoI) and Probabilism

Probabilism is the following norm:

(Prob) At any time in an agent's credal life, it ought to be the case that her credence function  $c$  at that time is a probability function. That is,

- $c(\top) = 1; c(\perp) = 0$ ;
- $c(X \vee Y) = c(X) + c(Y)$  for all exclusive  $X$  and  $Y$ .

Like (Prob<sub>2</sub>), (Prob) applies to an agent at any time during her credal life. Let (Prob<sup>sb</sup>) be the version of (Prob) that applies only to superbabies—that is, it demands only that an agent's *initial* credence function be a probability function. Then the relationship between (PoI) and (Prob<sup>sb</sup>) is straightforward: (PoI)  $\Rightarrow$  (Prob<sup>sb</sup>). After all, the proportion of worlds in which a tautology is true is 1, and the proportion in which a contradiction is true is 0; furthermore, the proportion of worlds in which  $X \vee Y$  is true is just the sum of the proportion in which  $X$  is true with the proportion in which  $Y$  is true, providing  $X$  and  $Y$  are exclusive. So  $c_0$  is a probability function. Thus, by justifying (PoI), we thereby justify (Prob<sup>sb</sup>).

### 6.2 (PoI) and Regularity

Regularity is the following norm (cf., for instance, [Kemeny, 1955]):

(Reg) At the beginning of an agent's credal life, it ought to be that her credence function  $c$  is regular. That is,

- For all  $X \not\equiv \perp$ ,  $c(X) > 0$ .

Again, (PoI)  $\Rightarrow$  (Reg). After all, if  $X \not\equiv \perp$ , then there is some world in  $\mathcal{W}_{\mathcal{F}}$  at which  $X$  is true. Thus, the proportion of worlds in which  $X$  is true is positive.

Thus, as well as an argument for the Principle of Indifference, the Argument from Accuracy gives a novel argument for Regularity as well. Some existing arguments for Regularity note that, if an agent violates it and updates by conditioning, there are certain propositions they could never come to learn. Thus, they attribute the irrationality of violating Regularity to the

trouble it stores up for a possible future. The Argument from Accuracy, by contrast, attributes this irrationality to the trouble it makes at the time the irregular initial credence function is held. The trouble in question is that such a credence function risks greater inaccuracy than it is necessary to risk.

### 6.3 (PoI) and the Principal Principle

The Principal Principle is a chance-credence norm [Lewis, 1980]. That is, it governs the relationship between your credences in propositions concerning chances and your credences in other propositions. Like David Lewis, I will take the Principal Principle to govern only superbabies. Consider such a superbaby. Suppose that amongst the propositions this agent considers are ones concerning what we shall call a *heady coin*. We say that a coin is *heady* iff the chance of the coin landing heads is either 60% or 90%. How strongly ought this agent to believe the following proposition: *The heady coin will land heads*? Lewis' Principal Principle says the following:

$$c(\text{Heads}) = (0.6 \times c(\text{Chance of heads is 60\%})) + (0.9 \times c(\text{Chance of heads is 90\%}))$$

More generally:

(PP) Let  $\mathcal{C}$  be the set of probability functions that might give the chances. For each  $ch$  in  $\mathcal{C}$ , let  $C_{ch}$  be the proposition *ch gives the chances*. Then, at the beginning of her credal life, an agent ought to have a credence function  $c$  such that

$$c(X) = \sum_{ch \in \mathcal{C}} c(C_{ch})ch(X)$$

for all  $X$  in the domain of the chance functions.<sup>11</sup>

That is, an agent's initial credence in a proposition ought to match her initial expectation of its chance.

This seems plausible. The problem is that (PP) and (PoI) are incompatible. Consider the example of the superbaby considering the heady coin:  $\mathcal{C} = \{ch_1, ch_2\}$ , where  $ch_1(\text{Heads}) = 0.6$  and  $ch_2(\text{Heads}) = 0.9$ , and

$$\mathcal{F} = \{C_{ch_1} \& \text{Heads}, C_{ch_1} \& \text{Tails}, C_{ch_2} \& \text{Heads}, C_{ch_2} \& \text{Tails}\}$$

Then (PoI) demands  $c(\text{Heads}) = 0.5$ , while (PP) demands  $0.6 \leq c(\text{Heads}) \leq 0.9$ . How, then can we resolve this tension between these two plausible principles?

---

<sup>11</sup>In fact, this is strictly weaker than Lewis' original version. It is Jenann Ismael's formulation restricted to the agent's initial credence functions [Ismael, 2008]. It will serve our purposes best here.



At this point, the Argument from Accuracy might provide a solution. I don't offer this as the only solution available; but, as we will see, it has plausible consequences. As we saw, it turns on (Minimax), which says:

(Minimax) Suppose  $o, o'$  in  $\mathcal{O}$  and

- (i)  $o$  worst-case dominates  $o'$
- (ii) There is no  $o''$  in  $\mathcal{O}$  such that  $o''$  worst-case dominates  $o$

Then  $o'$  is irrational.

In the Argument from Accuracy presented above, we took  $\mathcal{O}$  to be the set of all possible credence functions on the algebra  $\mathcal{F}$  of propositions about which the agent has an opinion. (Minimax) then says of every credence function except  $c_0$  that it is irrational. But, if there are other norms in play—such as the Principal Principle—perhaps we ought to restrict  $\mathcal{O}$  to the set of all credence functions that obey those other norms. Thus, we might first restrict to the credence functions that obey (PP); and then ask for the one *amongst those* that minimises its maximum inaccuracy. If  $c_0$  obeys (PP), then it is mandated. But if, as in the example,  $c_0$  does not obey (PP), then some other credence function is sanctioned. Let's see this in action in the example above.

For any credence function that satisfies (PP) in this situation, there is  $0 \leq x \leq 1$  such that

$$\begin{aligned} c(C_{ch_1}) &= x \\ c(C_{ch_2}) &= 1 - x \\ c(Heads) &= 0.6c(C_{ch_1}) + 0.9c(C_{ch_2}) = 0.9 - 0.3x \\ c(Tails) &= 0.4c(C_{ch_1}) + 0.1c(C_{ch_2}) = 0.1 + 0.3x \end{aligned}$$

We can then calculate the Brier score at each of the four possible worlds in  $\mathcal{W}_{\mathcal{F}}$ , namely,

$$w_1 = C_{ch_1} \& Heads, \quad w_2 = C_{ch_1} \& Tails, \quad w_3 = C_{ch_2} \& Heads, \quad w_4 = C_{ch_2} \& Tails$$

This gives:

$$\begin{aligned} B(c, w_1) &= 2(0.1 + 0.3x)^2 + 2(1 - x)^2 \\ B(c, w_2) &= 2(0.9 - 0.3x)^2 + 2(1 - x)^2 \\ B(c, w_3) &= 2(0.1 + 0.3x)^2 + 2x^2 \\ B(c, w_4) &= 2(0.9 - 0.3x)^2 + 2x^2 \end{aligned}$$

Then define the following credence function:

$$c^*(C_{ch_1}) = 0.5 \quad c^*(C_{ch_2}) = 0.5 \quad c^*(Heads) = 0.75 \quad c^*(Tails) = 0.25$$

Then, if  $c \neq c^*$  and  $c$  satisfies (PP), we have

$$\max_{1 \leq i \leq 4} B(c^*, w_i) < \max_{1 \leq i \leq 4} B(c, w_i)$$

Thus, if we apply Minimax to the set of credence functions that satisfy (PP) in this case, it says of every one of those credence functions except  $c^*$  that it is irrational. Thus, we ought to adopt  $c^*$  in that situation. That is, in the absence of evidence, we ought to be indifferent between the chance hypotheses, and then appeal to (PP) to set the other credences.

Interestingly, this result is sensitive to the composition of the set  $\mathcal{F}$ . Suppose  $\mathcal{F} = \{Heads, Tails\}$ . And suppose that, although she has no credences in  $C_{ch_1}$  and  $C_{ch_2}$ , our agent does nonetheless know that one or other is true. Now, our original formulation of (PP) applied only to agents who have credences in the chance hypotheses. But it seems reasonable to drop this requirement and say not that an agent ought to have an initial credence function that satisfies (PP), but that she ought to have an initial credence function *that can be extended to a credence function that satisfies (PP)*. This demand does not require her to have credences in the chance hypotheses. Thus, for an agent who knows that either  $C_{ch_1}$  or  $C_{ch_2}$  is true but for whom  $\mathcal{F} = \{Heads, Tails\}$ , she satisfies this more general version of (PP) iff there is  $0 \leq x \leq 1$  such that

$$\begin{aligned} c(Heads) &= 0.9 - 0.3x \\ c(Tails) &= 0.1 + 0.3x \end{aligned}$$

Now  $\mathcal{W}_{\mathcal{F}} = \{w_1 = Heads, w_2 = Tails\}$ . Thus, we have the following Brier scores:

$$\begin{aligned} B(c, w_1) &= 2(0.1 + 0.3x)^2 \\ B(c, w_2) &= 2(0.9 - 0.3x)^2 \end{aligned}$$

And the credence function that minimises its maximum inaccuracy is  $c^+$  defined as follows:

$$c^+(Heads) = 0.6 \quad c^+(Tails) = 0.4$$

That is,  $c^+ = ch_1$ . Thus, if  $\mathcal{F}$  includes chance hypotheses and we apply (Minimax) to the credence functions that satisfy (PP), we get that our credence in Heads oughts to be 0.75. On the other hand, if  $\mathcal{F}$  leaves out the chance hypotheses and we apply (Minimax) to the credence functions that can be extended to ones that satisfy (PP), our credence in Heads ought to be 0.6.

On reflection, this seems right. If we care only about the accuracy of our credences in the non-modal propositions, we ought to set those credences as near as possible to those demanded by  $c_0$ . In the example under

consideration, that means assigning all weight to  $C_{ch_1}$ —indeed, we set our credences equal to  $c^\dagger = ch_1$ . However, if we also care about the accuracy of our credences in the chance hypotheses, such a strong weighting towards  $C_{ch_1}$  is risky, since it will be maximally inaccurate if  $ch_2$  turns out to give the true chances and  $C_{ch_2}$  is true. Thus, we ought to assign each chance hypothesis equal credence and set our credences in Heads and Tails as (PP) demands—the resulting credence function is  $c^*$ .

However, this merely shifts the question, which now becomes: Should we care about the accuracy of our credence in the chance hypotheses? I would say that we should. They are, after all, simply claims about the way the world is, just like Heads and Tails. Just because they are modal claims concerning chance, this should not deprive the accuracy of credences in them of cognitive value. Thus, we should adopt  $c^*$ .

Interestingly, this gives a different solution to this problem from that given by Jon Williamson [Williamson, 2007]. For Williamson, we ought to adopt  $c^\dagger$ . For him, our credences in the non-modal propositions ought to be as close as possible to  $c_0$ , where distance is measured by cross-entropy. The preceding considerations tell against this view.

Williamson might respond as follows: We are only interested in chance hypotheses to the extent that they inform our credences in the non-modal propositions. So we are only interested in the accuracy of the chance hypotheses to the extent that it affects the accuracy of our credences in the non-modal propositions. Thus, we ought to seek the credence function over the non-modal propositions that minimises maximum inaccuracy. Perhaps, but I'm not convinced. Of course, it may be that, for pragmatic reasons, we are only interested in the non-modal facts—perhaps all of our choices are ultimately made on the basis of our credences in the non-modal propositions, so it is only their accuracy that has pragmatic value. But we are not interested here in the pragmatic value of accuracy beliefs. We are interested in their cognitive value. And then we have no reason to consider only the non-modal propositions and our credences in them.

## 7 The consistency of (PoI)

In the preceding sections, I have presented a novel justification of the Principle of Indifference. However, it might be objected that this putative norm is inconsistent and thus unjustifiable. In this section, I take up this objection.

To appreciate the alleged inconsistency, consider an agent at the beginning her credal life who entertains only two propositions: *Red* says that the handkerchief in my pocket is red; *Blue* says that it is blue; the agent knows that *Red* and *Blue* are exclusive and exhaustive. The Principle of Indifference demands of such an agent that she assign credence  $\frac{1}{2}$  to each proposi-

tion. However, suppose instead that she entertains credences in only three propositions, namely, *Red*, *Light Blue*, and *Dark Blue*, which she knows to be exclusive and exhaustive. Then the Principle of Indifference demands of her that she assign credence  $\frac{1}{3}$  to each proposition. Thus, the Principle of Indifference makes inconsistent demands: it demands that  $c(\text{Red}) = \frac{1}{2}$  and it demands that  $c(\text{Red}) = \frac{1}{3}$ . It is impossible to satisfy both demands. Thus, it is impossible to satisfy the Principle of Indifference. Thus, this principle is not a norm (by the meta-normative ought-can principle) and thus cannot be justified.

The problem with this objection is that it mistakes the content of the Principle of Indifference. The principle does not demand of *every* agent that she assign credence  $\frac{1}{2}$  to *Red* in the absence of evidence; nor does it demand of *every* agent that she assign credence  $\frac{1}{3}$  to *Red* in the absence of evidence. Rather, as was made clear in our statement of (PoI) above, the credence it demands an agent assign to *Red* in the absence of evidence depends on the set  $\mathcal{F}$  of propositions to which she assigns credences.

- If  $\mathcal{F} = \{\text{Red}, \text{Blue}\}$ , then  $c_0^{\mathcal{F}}(\text{Red}) = \frac{1}{2}$  and this is the credence that (PoI) demands of our agent.
- If  $\mathcal{F} = \{\text{Red}, \text{Light Blue}, \text{Dark Blue}\}$ , then  $c_0^{\mathcal{F}}(\text{Red}) = \frac{1}{3}$  and this is the credence that (PoI) demands.

Thus, there is no inconsistency.

The point is that the credence that the Principle of Indifference demands a superbaby assign to a particular proposition depends on the other propositions to which she assigns credences. At first, this may seem implausible, but I submit that it is entirely appropriate. After all, intuitively, the Principle of Indifference demands that we divide our credences equally over the possibilities. But which possibilities? Not all of the genuine metaphysical possibilities, since we cannot expect a superbaby to be aware of all of these—that would require *a posteriori* knowledge. Thus, the possibilities over which we demand that an agent divide her credences equally must be more limited or more coarse-grained than this. But we might nonetheless require that whatever these coarse-grained possibilities are, they must respect the ‘true symmetries’ of the situation in the way that one might think the possible worlds relative to  $\mathcal{F} = \{\text{Red}, \text{Blue}\}$  do, while the possible worlds relative to  $\mathcal{F} = \{\text{Red}, \text{Light Blue}, \text{Dark Blue}\}$  do not. Again, however, to demand that is to demand too much. After all, the true symmetries of the situation—like the genuine metaphysical possibilities—can only be known *a posteriori*. So we cannot demand of a superbaby that she set her credences in accordance with them. Instead, we demand that a superbaby divide her credence equally over the possibilities *gained as finely as the propositions she entertains will allow*: that is, a possibility in this context is way of consistently assigning truth values to all the propositions that the

agent entertains; in other words, it is a possible world relative to the set  $\mathcal{F}$  of propositions she entertains; that is, it is an element of  $\mathcal{W}_{\mathcal{F}}$ . So of course if an agent entertains *Red*, *Light Blue*, and *Dark Blue*, there are three possibilities over which she must divide her credences; but if she entertains only *Red* and *Blue*, there are just two. So the Principle of Indifference will make different demands on her depending on the set of propositions she entertains.

## 8 Conclusion

We have presented a novel argument for the Principle of Indifference. The Argument from Accuracy says that an agent who fails to divide their credences equally over the possibilities is irrational because she risks greater inaccuracy than she needs to. The argument does not appeal to any measure of evidential support; nor does it require us to say what is so bad about going beyond the information contained in our evidence. Moreover, it is a non-pragmatic argument: it appeals only to the cognitive value of credal states; it does not appeal to the pragmatic values of the choices that a credal state might lead us to make. In these ways, it avoids the problems with existing arguments for the Principle of Indifference.

The present formulation of the Principle of Indifference and the justification provided are restricted to agents with credences in only finitely many propositions. What happens in the infinite case? This is far from clear. Of course, we know that the most natural formulation of the Principle of Indifference is incompatible with Countable Additivity: If  $A_1, A_2, \dots$  is a countable partition, the Principle of Indifference demands that  $c(A_1) = c(A_2) = \dots$ , while Countable Additivity demands that  $\sum_i c(A_i) = 1$ ; and it is easy to see that these demands cannot both be satisfied. Furthermore, Bertrand's paradoxes [Bertrand, 1889] and van Fraassen's cube factory example [van Fraassen, 1989] pose problems for the Principle of Indifference on continuous probability spaces. What's more, it is no longer so straightforward to define inaccuracy measures for credence functions over infinitely many propositions. The definition of the Brier score does not extend naturally to infinite sets. How we might adapt the Argument from Accuracy to shed light on these puzzles is work for the future.

## References

- [Armendt, 1993] Armendt, B. (1993). Dutch Books, Additivity, and Utility Theory. *Philosophical Topics*, 21(1).
- [Bertrand, 1889] Bertrand, J. (1889). *Calcul de probabilités*. Gauthier-Villars, Paris.
- [Buchak, 2013] Buchak, L. (2013). *Risk and Rationality*. Oxford University Press.

- [Christensen, 1996] Christensen, D. (1996). Dutch-Book Arguments Depragmatized: Epistemic Consistency for Partial Believers. *The Journal of Philosophy*, 93(9):450–479.
- [Easwaran, 2013] Easwaran, K. (2013). Expected Accuracy Supports Conditionalization - and Conglomerability and Reflection. *Philosophy of Science*, 80(1):119–142.
- [Easwaran, ms] Easwaran, K. (ms). Dr. Truthlove, or, How I Learned to Stop Worrying and Love Bayesian Probability. <http://dl.dropboxusercontent.com/u/10561191/Unpublished/Truthlove.pdf>.
- [Goldman, 2002] Goldman, A. (2002). The Unity of the Epistemic Virtues. In *Paths to Knowledge: Private and Public*. Oxford University Press, New York.
- [Hájek, 2008] Hájek, A. (2008). Arguments For—Or Against—Probabilism? *The British Journal for the Philosophy of Science*, 59(4):793–819.
- [Hájek, ms] Hájek, A. (ms). Staying Regular. <http://hplms.berkeley.edu/HajekStayingRegular.pdf>.
- [Ismael, 2008] Ismael, J. (2008). Raid! Dissolving the Big, Bad Bug. *Noûs*, 42(2):292–307.
- [James, 1905] James, W. (1905). The Will to Believe. In *The Will to Believe, and other essays in popular philosophy*. Longmans Green, New York.
- [Jaynes, 1957] Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, 106(4):620–630.
- [Joyce, 1998] Joyce, J. M. (1998). A Nonpragmatic Vindication of Probabilism. *Philosophy of Science*, 65(4):575–603.
- [Joyce, 2009] Joyce, J. M. (2009). Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief. In Huber, F. and Schmidt-Petri, C., editors, *Degrees of Belief*. Springer.
- [Kemeny, 1955] Kemeny, J. G. (1955). Fair Bets and Inductive Probabilities. *Journal of Symbolic Logic*, 20(3):263–273.
- [Lewis, 1980] Lewis, D. (1980). A Subjectivist’s Guide to Objective Chance. In Jeffrey, R. C., editor, *Studies in Inductive Logic and Probability*, volume II. University of California Press, Berkeley.
- [Paris and Vencovská, 1990] Paris, J. B. and Vencovská, A. (1990). A note on the inevitability of maximum entropy. *International Journal of Approximate Reasoning*, 4:181–223.
- [Pettigrew, 2013] Pettigrew, R. (2013). A New Epistemic Utility Argument for the Principal Principle. *Episteme*.
- [Predd et al., 2009] Predd, J., Seiringer, R., Lieb, E. H., Osherson, D., Poor, V., and Kulkarni, S. (2009). Probabilistic Coherence and Proper Scoring Rules. *IEEE Transactions on Information Theory*, 55(10):4786–4792.
- [Rawls, 1975] Rawls, J. (1975). *A Theory of Justice (revised edition)*. Oxford University Press, New York.

- [Savage, 1951] Savage, L. J. (1951). The Theory of Statistical Decision. *Journal of the American Statistical Association*, 46(253):55–67.
- [Skyrms, 1984] Skyrms, B. (1984). *Pragmatics and Empiricism*. Yale University Press, New Haven.
- [van Fraassen, 1989] van Fraassen, B. C. (1989). *Laws and Symmetry*. Oxford University Press, Oxford.
- [Vineberg, 2001] Vineberg, S. (2001). The Notion of Consistency for Partial Belief. *Philosophical Studies*, 102:281–296.
- [White, 2009] White, R. (2009). Evidential Symmetry and Mushy Credence. *Oxford Studies in Epistemology*, 3:161–186.
- [Williamson, 2007] Williamson, J. (2007). Motivating Objective Bayesianism: From Empirical Constraints to Objective Probabilities. In Harper, W. L. and Wheeler, G. R., editors, *Probability and Inference: Essays in Honor of Henry E. Kyburg Jr.*, pages 155–183. College Publications, London.